



TITLE:

The predictive capacity of biological theories
(Proceedings of the CAPE International
Workshops, 2012. Part I: IHPST, Paris - CAPE,
Kyoto philosophy of biology workshop)

AUTHOR(S):

Leconte, Gauvain

CITATION:

Leconte, Gauvain. The predictive capacity of biological theories (Proceedings of the CAPE International Workshops, 2012. Part I: IHPST, Paris - CAPE, Kyoto philosophy of biology workshop). CAPE Studies in Applied Philosophy and Ethics Series 2013, 1: 60-68

ISSUE DATE:

2013-02-12

URL:

<https://doi.org/10.17983/203248>

RIGHT:

The predictive capacity of biological theories

Gauvain Leconte

Perhaps it is time to investigate the possibility that biology has no empirical laws of evolution because of the strategies of model building that biologists have adopted.

E. Sober (1997)

There have been intense debates since the 1950's about the possibility to use the theory of natural selection to derive predictions. Most of the discussions about the predictiveness of biology (the “predictiveness thesis”) involved arguments about the status of biological theories, the presence of laws in biology, and so on. But few comments have been made on the different kinds of predictions and on the specific status of predictions in biology. I focus on this issue, and maintain that it is not necessary to claim that biological and physical theories have identical status and realize predictions in the same way to support the predictiveness thesis.

Some precisions about predictions

There is a standard model, or family of models, of predictions: the *covering-law models* of prediction. The most famous of these models is Hempel & Oppenheim (1948)'s deductive-nomological model, where a statement E is deduced from lawlike statements (universal proposition) and statements about initial conditions.

The DN model and the other covering-law models built by Hempel where laws are statistical or the statement E induced and not deduced, involves symmetry between explanation and prediction. Prediction and explanation being in all these models inference of an event E from a set of laws and a set of initial conditions, the only difference is not in the logical form of the prediction or explanation but in the knowledge of E: if it is known before the deduction, it is an explanation, if it is not, it is

a prediction.

Moreover, in these models all the laws used to predict a phenomenon E come from the same theory. These statements are all formulated in a language with the same domain of interpretation. No laws from other theories seem to be necessary to the derivation of predictions.

For the standard model, a prediction is thus a *deduction* of a *future* or *unknown* event from initial conditions and laws belonging to *one theory*.

If the covering-laws models of explanation have been criticized and alternative models proposed since the 1960's, it is only recently, with the debate around the role of novel predictions as a proof for scientific realism, that new accounts of prediction have been proposed.

It is now for example widely acknowledged that a prediction does not have to be a forecast, i.e. a statement about the future, or a statement about something *unknown* in the past, present or future. This kind of temporal account of prediction can lead to severe paradoxes: the same inference of an event would be an explanation or a prediction for different scientists depending on the information they knew at the time they made this inference. But it seems that what is relevant is not the information they *knew* but rather the information they *used* to infer this event. Worrall (forthcoming) has thus argued that there is no reason to accept that only because a fact is unknown at the time he is inferred from a hypothesis it will count as a predicted fact, and count more than a fact already known (an accommodated fact): a phenomenon E is predicted by a hypothesis if it has not been used in the construction of this hypothesis, i.e in the setting of the free parameters of the hypothesis. It is the *heuristic* novelty and not the *temporal* novelty of the fact which matters. The relevant distinction is then not between prediction and explanation, but between *ad hoc* and non *ad hoc* entailment of a fact by a theory.

Moreover: we do not need *laws* to make predictions. The rise of statistical methodologies in all the fields of science has given birth to other predictive processes than deductions in a mathematical framework. Forster and Sober (1997, 17) showed that predictions do not have to be made in a deductive framework to distinguish genuine predictions from *ad hoc* accommodations:

Although we reject the hypothetico-deductive picture of science, we do accept the usual conclusion that there is an important distinction to be drawn between reasonable revision and *ad hoc* evasion.

It is then possible to talk about predictions even when scientists do not deduce their predictions from laws but draw their predictive models from statistical databases.

Finally, one of the facts often overlooked is the importance of a *plurality of hypothesis* for predictions: in order to predict a concrete event it is often needed to gather a lot of information about the initial conditions of the system, and it is really extraordinary if this information involve phenomenon described by one theory only. Prediction is an activity for those Sober (1997, 461) calls “pluralists”, the scientists who consider phenomena on the border of different theories.

My purpose is not to build a non-standard model or an alternative to the standard model, but only to stress the diversity of predictions and of processes of prediction-making in science. Too often it is considered that the covering-law models give a fair account of prediction and that troubles begin with the thesis of the symmetry of prediction and explanation. That is precisely the case of the controversy about the predictiveness of biology, which I evoke briefly in the following section.

History of the problem of the predictiveness of biology

It has soon been remarked that the symmetry between explanation and prediction of the DN model and its followers do not seem applicable to biological theories. Scriven (1959) argues that the selection theory relies on the concept of “fitness” which refers to a connection of an organism to his environment. Thus we cannot predict the evolution of the fitness of an organism nor the survival of its offspring because we cannot predict the environmental changes. But we can, *a posteriori*, give an explanation for the evolutionary success or extinction of an organism. Mayr (1961, 150) explicitly subscribes to Scriven’s argument:

Probably nothing in biology is less predictable than the future course of evolution. Looking

at the Permian reptiles, who would predicted that most of the more flourishing groups would become extinct and that one of the most undistinguished branches would give rise to the mammals?

In other words, even if they had been really good in biology, dinosaurs could not have predicted their own extinction (but we can explain it).

All these critics of the symmetry between explanation and prediction accept the description of predictions of the covering-law model¹. Several authors like Mary B. Williams or Zdislaw Kochanski have argued for the predictiveness of biology against the conclusion of Mayr and Scriven, but without explicitly questioning the standard model of prediction. They aimed at revealing implicit laws and deduction implicitly used by the students of natural selection.

This view reached its height with Mary B. Williams' (1970, 1973b) axiomatic of the theory of natural selection and quantitative derivation of predictions from axioms.

For Williams, if we can axiomatize statements of the theory of evolution in order to deduce predictions about *populations* of entities like we do in statistical physics for *populations* of molecules, the apparent randomness of biological phenomena is no obstacle to the predictiveness of the theory of natural selection. The randomness only come when we want to state which particular individual in those populations have or not the feature stated in the prediction: “the logical peculiarity [of evolutionary prediction] is not a property of evolutionary theory, it is a property of the human sized perspective on evolutionary theory” (Williams 1973a, 536).

Although I share the predictiveness thesis with Williams and Zochanski, I think that their view and strategy are misled and may even undermine the predictiveness thesis. Without criticizing the standard model of prediction it is not possible to give a fair account of how predictions and models are built in evolutionary biology. I expose in the following section a case of biological prediction to show how it would seem unnatural to analyze it with the standard model of predictions.

A case of biological prediction: Can we predict the *loci* of evolution?

Different studies in microbiology showed that on the long-term (2 decades, 30 000

generations) populations of bacteria such as *Escherichia coli* (Lenski 2004) or bacteriophage viruses (Wichman et al. 1999) drawn from the same ancestral population and under the same selective pressure, have known parallel evolution, both at phenotypic and genotypic levels.

This has lead to the extensive study of connections between what Mayr called functional and evolutionary biology and to the following question: can we predict, for a given selective pressure, what will be the transformation of the code which will lead to adaptive mutations?

Two biologists of development, Orgogozo and Stern (2008, 2009) have studied the hypothesis that most of the evolutionary relevant mutations take place in the cis-regulatory regions of genes² to claim that “genetic evolution is predictable”.

Orgogozo and Stern have built a database of published studies providing compelling evidence for the individual genetic mutations causing phenotypic variation. This methodology is clearly based on statistical analysis: the statistical tests used by Orgogozo and Stern are G test, a measure of likelihood ratio often used in biological sciences since the 1980s.

Stern and Orgogozo do not mention or use any “law”. They do not try to generalize the *cis-regulatory mutations* hypothesis but to prove that without empirical data analysis, this hypothesis has no predictive power. Thus it seems natural to consider that Stern and Orgogozo’s work does not fit in a hypothetico-deductive picture of science or in the standard model of prediction.

Moreover, Orgogozo and Stern precise their prediction by appealing to other biological sciences: developmental biology for instance is used to prove that some genes have more chance to have specific effects than other genes in the same regulatory network. For example, the development of the bristles, sensory organs on the back of *Drosophila melanogaster*, involves several genes but one, *scute*, is an “input-output device”: during the development it integrates an extensive array of inputs, the regulatory state, and produce an on or off transcriptional output. It is then responsible for the differentiation of almost all cells of the back of the fly. That is why such genes like *scute* have more chance to be involved in long term evolution: they are what Orgogozo and Stern call “hot spot genes”. This is the basis for the following prediction: “We predict

that the *cis*-regulatory region of input-output genes may be hotspots for other phenotypic characteristics” (Orgogozo and Stern 2009, 747)

It seems obvious to say that Orgogozo and Stern are *pluralists*: they merge models from different theories of the biological sciences to realize predictions. The conclusion of their paper is clear on the subject: “By fusing developmental and evolutionary genetics, evolutionary biologists may be able to predict, in a probabilistic sense, the mutations underlying phenotypic evolution” (Orgogozo and Stern 2008, 2171)

Finally, because this prediction refers to long term evolution, it has not been today used to predict unknown mutations: “our prediction based on regulatory network is a *post hoc* explanation” (Ibid., 2170). But it does not stop Orgogozo and Stern to use the word “prediction” because they were able to infer the evolutionary role of mutations in the *cis*-regulatory region of *scute* with only a pattern of developmental regulation, i.e. by using considerations about the ontogeny of *Drosophila melanogaster* and not about its phylogeny. This prediction is then *heuristic* novel.

The study of Orgogozo and Stern's work then presented a biological prediction relative to evolution far from fitting naturally in the standard model of prediction. But if we acknowledge that biological predictions do not fit in covering-law models, can we still support a *strong* predictiveness thesis claiming not only that there are predictions in biology but also that biology and physics have the same predictive power? The last section deals with this problem.

The predictive capacity of the theory of natural selection

We must distinguish between:

- Predictive process: how a statement about phenomena or facts results, in a non *ad hoc* way, from a hypothesis or a theory connected with auxiliaries assumption through a series of inferences.
- Predictive capacity (of a theory): the range of phenomena predicted by a theory. Roughly speaking, it refers to the diversity of phenomena whose probability become higher if a theory or hypothesis is assumed as correct, i.e. the phenomena E such as $P(E|T) > P(E)$.

However, it is difficult to compute directly this increase of probability because we do not know the prior probability $P(E)$: all we can do is to compare its probability within two different theoretical backgrounds ($P(E|T)$ and $P(E|T')$). If a phenomenon had a low probability for all theories before being predicted by T , we can be qualified as a *surprising prediction*. Surprising predictions are thus symptomatic of an important predictive capacity.

Standard and non standard model of prediction describe *predictive processes*, but what is at stake when we ask if biology and physics have the same predictive power is *predictive capacity*. Two theories may be used in different predictive processes (in fact each theory offer different possibilities of prediction-making) but have the same predictive capacity. We can say that prediction are different in biology and in physics but still support the thesis that biological theories and physical theories have similar predictive capacities.

However we can only evaluate the predictive capacity of a theory if we can identify clearly which derivations of empirical facts count as prediction. Popper for example suspects “Darwinism” of *ad hoc* evasion: “Darwin’s theory of sexual selection is partly an attempt to explain falsifying instances of this theory; such things, for example, as the peacock’s tail, or the stag’s antlers” (Popper 1976, 199) But if we adopt a heuristic definition of prediction and accommodation, Darwin’s theory of sexual selection appears not as an *ad hoc* evasion but as a prediction. Even if we had known nothing of the peacock’s tail Darwin’s theory of selection is sufficient to predict the selection of traits without any survival’s utility other than in the process of mating.

In this nonstandard account of prediction, Ramsus Grønfeldt Winther (2009, 894), has shown that there are good reasons to think that parallel evolution is a *surprising* prediction of the theory of natural selection, because it could not have been explained by evolutionary theories with no reference to selection and selective pressure. In the same manner the prediction of Orgogozo and Stern is a surprising one, because the importance of *cis*-regulatory mutations of hot spot genes over long-term evolution is a consequence of selective pressure on mutations. And if it is confirmed it could be counted as an important corroboration of the theory of natural selection.

It is clear that evolutionary biology does not have the same level of precision than

general relativity of quantum mechanics. But no other contemporary theory to my knowledge has such an extension over all the phenomena of a discipline. Everyone, including the opponent to the predictiveness thesis recognizes that one of the major theoretical virtues of Darwinism is its unification of a large range of phenomena. It compensates its inferiority of predictive precision by a superiority of diversity of phenomena.

Conclusion

We may never be able to fully measure the predictive capacity of a theory. It is maybe more a matter of scientist's intuition than a matter of crude computation. However, by distinguishing predictive capacity and predictive process we can, instead of opposing predictive capacity and explanatory unification as two different theoretical virtues, consider that they reveal the same virtue. We can then defend the status of the selectionist theory of evolution as a mature predictive theory without identifying its predictions to the one found in physics.

Notes

1. Recently the debate over the symmetry of prediction and explanation has reappeared in the field of methodology of historical natural science (c.f., Cleland 2011). Cleland uses Scriven's argument and if she criticizes the account of explanation of covering-law model it is clear that for her this model is a sound way to describe predictions.
2. All genes are composed of two fundamentally different regions: a region encoding the gene product – a protein or RNA – and adjacent *cis*-regulatory DNA that encodes the instructions governing when and where the gene product will be produced.

References

- Cleland, C. E. 2011. Prediction and Explanation in Historical Natural Science. *British Journal for the Philosophy of Science* 62(3): 551–582.
- Forster, M. and Sober, E. 1994. How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions. *British Journal for the Philosophy of Science* 45(1): 1–35.

- Hempel, C. G., and Oppenheim, P. 1948. Studies in the Logic of Explanation. *Philosophy of Science* 15(2): 135–175.
- Kochanski, Z. 1973. Conditions and Limitations of Prediction-Making in Biology. *Philosophy of Science* 40(1): 29–51.
- Lenski, R. E. 2004. Phenotypic and Genomic Evolution during a 20,000-generation Experiment with the Bacterium *Escherichia coli*. *Plant Breeding Reviews* 24(2): 225–266.
- Mayr, E. 1961. Cause and Effect in Biology: Kinds of Causes, Predictability, and Teleology are Viewed by a Practicing Biologist. *Science* 134: 1501–1506.
- Orgogozo, V. and Stern, D. 2008. The Loci of Evolution: How Predictable Is Genetic Evolution? *Evolution* 62(9): 2155–2177.
- . 2009. Is Genetic Evolution Predictable? *Science* 323(5915): 746–751.
- Popper, K. 1976. *Unended Quest: An Intellectual Autobiography*. 2nd edition. London: Routledge.
- Scriven, M. 1959. Explanation and Prediction in Evolutionary Theory. *Science* 130:477–482.
- Sober, E. 1997. Two Outbreaks of Lawlessness in Recent Philosophy of Biology. *Philosophy of Science* 64: S458–S467.
- Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M., and Bull, J. J. 1999. Different Trajectories of Parallel Evolution during Viral Adaptation. *Science* 285(5426): 422–424.
- Williams, M. B. 1970. Deducing the Consequences of Evolution: A Mathematical model. *Journal of Theoretical Biology* 29(3): 343–385.
- . 1973a. Falsifiable Predictions of Evolutionary Theory. *Philosophy of Science* 40(4): 518–537.
- . 1973b. The Logical Status of the Theory of Natural Selection and Other Evolutionary Controversies. In M. Bunge (Ed.) *The Methodological Unity of Science* (pp. 343–385). London: Springer.
- . 1982. The Importance of Prediction Testing in Evolutionary Biology. *Erkenntnis* 17(3): 291–306.
- Winther, R. G. 2009. Prediction in Selectionist Evolutionary Theory. *Philosophy of Science* 76: 889–901.